



**CON-TEXTOS KANTIANOS.**  
**International Journal of Philosophy**  
 N.º 6, Diciembre 2017, pp. 256-271  
 ISSN: 2386-7655  
 Doi: 10.5281/zenodo.1095821

## **A Guide to Kant's Treatment of Grace**

### *Una guía del tratamiento kantiano de la gracia*

PABLO MUCHNIK<sup>1</sup> & LAWRENCE PASTERNAK<sup>2</sup>

Emerson College, USA & Oklahoma State University, USA

#### **Abstract**

This Guide is designed to restore the theological background that informs Kant's treatment of grace in *Religion* to its rightful place. This background is essential not only to understand the nature of Kant's overall project in this book, namely, to determine the "association" or "union" between Christianity (as a historical faith) and rational religion, but also to dispel the impression of "internal contradictions" and "conundrums" that contemporary interpreters associate with Kant's treatment of grace and moral regeneration. That impression, we argue, is the result of entrenched interpretative habits that can be traced back to Karl Barth's reading of the text. Once we realize that such a reading rests on a mistake, much of the anxiety and confusion that plague current discussions on these issues can be put to rest.

#### **Key Words**

Grace, sanctifying, justification, original sin, total depravity, conundrum, Augustine, Calvin, Luther, Pietism

*Religion within the Boundaries of Mere Reason* is among Kant's most misunderstood and maligned texts. Both his critics and defenders alike raise worries about its alleged employment of an "Augustinian" conception of moral evil—a conception which in turn shapes Kant's views on grace and moral regeneration and is supposed to be responsible for the "wobbles," "conundrums," and "internal contradictions" associated with them.

In April 2017, the North American Kant Society, as part of the Central Division meeting of the American Philosophical Association, hosted a symposium to shed light on

---

<sup>1</sup> Associate Professor of Philosophy at Emerson College, USA. Email contact: [pablo\\_muchnik@emerson.edu](mailto:pablo_muchnik@emerson.edu).

<sup>2</sup> Professor of Philosophy at Oklahoma State University, USA. Email contact: [l.pasternack@okstate.edu](mailto:l.pasternack@okstate.edu).

these issues. The papers by Jacqueline Mariña, Robert Gressis, and Dennis Vanden Auweelle included in this *dossier* were part of that event.

Our goal in this introduction is to provide some helpful background to the contemporary debate, a reader's guide so-to-speak to this particularly baffling topic. We begin with a brief discussion of the overall project of the *Religion*, necessary to place the question of grace within its proper context (§1). We then proceed to analyze some of the key theological notions informing Kant's treatment of grace (§2). These notions play an important role in the "conundrum" criticism and will help us show that this criticism rests on a mistake (§3). We conclude by examining how the papers in the symposium may serve as harbingers of a post-conundrum era (§4).

### §1. How to read 'Religion': The Hermeneutics of Kant's Religious Hermeneutics

Although Kant originally planned to write a series of separate essays on religion for the *Berlinische Monatsschrift*, political circumstances led him to combine them into a single volume. Even though we do not know how much of the remaining three pieces Kant had already planned when writing the first, we do know from his correspondence in the year preceding the publication of *Religion* that his intention, from the start, was to write a series of essays exploring the scope of overlap between rational religion and Christianity (e.g. AA 11:358, 11:429). Hence, despite the fact that Kant originally planned to publish a series of self-standing essays, they were nonetheless all part of an ongoing program to explore the relationship between rational religion and Christianity as a historical faith.<sup>1</sup> Indeed, the project articulated in the correspondence is not only reflected throughout *Religion*'s four parts, but is also explicitly stated in its two prefaces.

In the First Preface, after a discussion of the Highest Good as the principle by which "morality inevitably leads to religion" (AA 6:6, 6:8n), Kant turns to the respective roles of philosophical and biblical theology. That discussion then leads to a description of *Religion* as an "experiment" [*Versuch*] (6:10) whose purpose is to determine the "association" or "union" [*Vereinigung*] which may exist between biblical theology and the "pure philosophical doctrine of religion" (AA 6:10). The Second Preface builds on this idea and presents it as key to understanding the title of the work itself.<sup>3</sup>

In response to concerns raised about the meaning and "intention hidden behind" (AA 6:12) its title, Kant explains that *Religion within [Innerhalb] the Boundaries [Grenzen] of Mere Reason* is meant to express an embedding relationship between the wider sphere of "historical faith" and a narrower sphere of natural or rational religion.

---

<sup>1</sup> The initial essay on evil was submitted to the *Berlinische Monatsschrift* in February 1792 and published in the April issue. The second essay was then submitted in the weeks following the publication of the first (AA 11:343). However, as Kant learned in June that it was denied an official imprimatur, he completed the remaining essays as one single volume, which was then submitted roughly two months later (AA 11:358). All four essays were thus completed within about a six-month period.

<sup>3</sup> While many interpreters have held that the *Religion* is guided by two experiments, there are good textual grounds to think otherwise. For reasons discussed in Pasternack 2014 and 2017, we use "experiment" in the singular.

Hence, just as the First Preface more tersely presents the *Religion* as an “experiment” [*Versuch*] (AA 6:10, 6:12) whose purpose is to determine the “association” or “union” [*Vereinigung*] between traditional and rational religion, so Kant now, after clarifying the title’s meaning, uses the imagery of two embedded domains to describe the nature of his “experiment.”

Kant’s method can be divided into three steps: identify a particular ‘fragment’ of historical faith which is to be evaluated, hold that fragment “up to moral concepts,” and, finally, determine whether or not the fragment can “lead back” to the “pure rational system of religion” (AA 6:12). In other words, Kant’s experiment consists in identifying a particular Christian doctrine for consideration, examining it in light of various moral concepts, and if these concepts are aptly reflected in the doctrine, determining its “unity” or “compatibility” with the “pure rational system of religion” (AA 6:12).

We will illustrate Kant’s procedure with various examples as we move forward, but first, we would like to note that the experiment of the *Religion* does not merely compare historical faith and rational religion in order to catalog their points of “unity”—it is also *testing* or *evaluating* Christian doctrine by way of this comparison. Although a few years later Kant responded to a royal rescript by claiming that *Religion* “make[s] no appraisal of Christianity” (AA 7:8), it is undeniable that the text is replete with evaluative language. For example, both in the Second Preface and throughout the book, Kant presents the religion of reason as “genuine [*eigentliche*] religion” (AA 6:12, see also: 6:112, 6:123n, 6:153) and contrasts it with mere “cult.” The suggestion is clear: to the extent that a historical doctrine coheres with rational religion, it is part of “genuine religion”; to the extent it falls outside those boundaries, it is either superfluous or dangerous, since it belongs to a “religion of rogation” (empty ritual) whose aim is to curry favor with God as substitute for our moral labor. In fact, Kant compares the two religious spheres with “oil and water” and expects that through his experiment the “purely moral religion (the religion of reason) [will] float to the top” (AA 6:13).

This evaluative dimension, however, does not entail that Kant “reduces” religion to morality, biblical to philosophical theology, or faith to reason—what it means is that for Kant the primary value of religious doctrines lies in their contribution to our moral edification (AA 6:48). When divorced from our moral improvement, the historical cognition of the Bible “belongs among the *adiaphora*” (AA 6:44n.) and can be treated as one sees fit.<sup>4</sup>

A few preliminary examples will make Kant’s comparative/evaluative procedure clear. Consider, for instance, the investigation of Original Sin throughout Part One. Kant assesses the broadly Augustinian tenets which comprise this doctrine, including that our moral status “comes to us by way of *inheritance* from our first parents” (AA 6:40). Even if, to our secular ears, this may sound preposterous, it was nonetheless part of the dominant

<sup>4</sup> Kant thus recognizes that there is much more to religion as historical faith than what of it overlaps with rational religion. Moreover, he even allows not only that there may be revelation and miracles, but also that they may have meaning beyond their moral significance (6:84, 6:105).

theological worldview at the time. What, then, does Kant conclude on this matter? Much as one would expect: he finds it “most inappropriate,” primarily on the ground that our moral status must be a “determination of the power of choice” (AA 6:39), not something acquired through the deed of another.

“Radical evil,” i.e., the idea that behind our particular misdeeds there is a singular root source of moral corruption, undergoes a similar examination. Many interpreters suppose that this notion is of Kant's own coinage, and so treat it as a philosophical (perhaps even a *transcendental*) thesis. Yet “radical evil” [radikal böse] comes from the Vulgate's *radix malorum* (1 Timothy 6:10), and is a central issue of Christian theology. Hence, as with other particular tenets bound up with the Augustinian conception of Original Sin, Kant explores the relationship of this theological notion to moral principles in order to determine what ‘compatibility’ or ‘unity’ it may have with rational religion.

Unlike inheritance, however, radical evil is a theological tenet that receives a more favorable assessment. Although Kant rejects its association with cupidity (AA 6:34) along with the Augustinian thesis of a “*corruption*” of our faculties (AA 6:35), he finds that this notion, when compared against moral concepts, can cohere with rational religion provided it is understood in kind, i.e., as “the presence in the subject of a common ground, itself a maxim, of all particular morally evil maxims” (AA 6:20).

What these examples show is that *Religion* should not be read as a linear philosophical argument. It is, instead, the result of a comparative analysis, one which considers a series of traditional religious doctrines, holds them (and their constituting tenets) up to moral principles, and determines, on that basis, their compatibility (or lack thereof) with rational religion. It would be a mistake, accordingly, to read Part One as an attempt to prove that humanity is universally evil. It should be read, rather, as an inquiry into how this item of historical faith fares when examined in light of the moral core of rational religion.<sup>5</sup> The same applies to matters of redemption in Part II, the role of the Church (Part III), and “ecclesiastical service” and “statutory ordinances” of Christianity (Part IV).

Unfortunately, readers of the *Religion* nowadays are often unfamiliar with its underlying theological concerns, and hence tend to confuse a candidate for evaluation with a straightforward philosophical assertion. This confusion leads one to misconstrue the text's inner dialectic, to overlook the fact that Kant is navigating through various Christian doctrines—some specific to Lutheran Pietism, some concerning debates between Lutheran Pietism and orthodox Lutheranism, some part of the broader Augustinian tradition. Our intent here is to make these theological underpinnings more explicit, not with an eye to present Kant as a hero of Christian Apologetics (Firestone/Jacobs 2008), but rather to appreciate his own interests in theology and his efforts as a “religious reformer” (Hare 2017, p. 72).

---

<sup>5</sup> Read this way, much of the contemporary scholarly anxiety about the missing “formal proof” of the universal propensity for evil can dissipate. Such anxiety is more an expression of old philosophical habits of reading than a real lacuna in Kant's own argument.

Furthermore, we wish to emphasize the importance of reading *all four parts* of *Religion* as stages in a single overarching project. Most contemporary readers are drawn exclusively to Part One and pay little heed to the rest. Such an approach, however, cannot be justified in light of the text's history: although Kant first planned to publish each essay as a separate unit, they were conceived as a four-part series meant to be read as a unitary project. To drop *Religion* after reading Part One is like dropping the first *Critique* after the Transcendental Aesthetic: the picture we will get is drastically distorted. Kant's grim pessimistic anthropology (Muchnik 2009, p. xxiii; Frierson 2010, p. 48-55; Velkeley 2014, p. 244) is only the starting-point of a lengthy inquiry into our moral regeneration, not its final word.

What we are suggesting, in sum, is to break loose from entrenched interpretative habits: we are inviting readers to see *Religion* as a coherent whole, a project shaped by a comparative and evaluative agenda that requires familiarity with the various Christian doctrines Kant examines. We propose that, to properly interpret the *Religion*, one must be able to recognize the particular doctrinal tenets with which Kant is engaging (in each Part and passage), recognize the philosophical principles by which they are assessed, ascertain what elements of historical faith meet Kant's evaluative standards, and (in light of the *parerga* with which each Part concludes) determine the limits of what rational religion can and cannot settle. To assist those interested in thinking of *Religion* in this fashion, we now offer a brief overview of the Christian conceptions of Original Sin and grace.

## §2. What is Grace?

Christian theology contains an array of different conceptions of (and roles for) grace, denominational disagreements, orthodoxies and heresies, and a number of famous historical debates (Augustine-Pelagius, Luther-Erasmus, Calvin-Arminius, etc.). However, 'grace' (*gracia*, *Gnade*) can be most generally understood as something God does on our behalf. This may be a specific act of divine aid, the endowment of some capacity, or a "forensic" judgment whereby we are relieved of some burden or forgiven for some transgression. Needless to say, we cannot possibly survey here the entire gambit of views existing on these matters. Our goal is more modest: to explain the specific conceptions of grace Kant *actually* considers.

First, Kant occasionally appeals to the Leibnizian distinction between the Kingdoms of Nature and Grace when discussing the Highest Good. For example, in the First *Critique*'s Canon of Pure Reason, he associates the postulate of immortality with the Kingdom of Grace, a realm in which happiness will be "distributed precisely in accordance with morality" (KrV A811/B839). While 'grace' in this context refers to God's agency (i.e., the distribution of happiness in our "future life"), insofar as happiness is allocated "precisely in accordance with morality" there is no overt forgiveness or mercy involved in the process. What matters is our "worthiness to be happy" (KrV A813/B841), and happiness is to be exactly apportioned according to that distributive criterion (KrV

A814/B842). Yet, Kant calls the “moral world” of our “future life” wherein the Highest Good is to obtain a *regnum gratiae* (A815/B843).

Second, in the Second *Critique*'s discussion of the postulate of immortality, Kant qualifies the Highest Good's principle of an “exact” correlation between morality and happiness, noting that we will never be “adequate to God's will” (AA 5:123) and hence are in need of “indulgence or dispensation, which do not harmonize with justice” (AA 5:123). Such accommodation looks different in the *Religion*, but the two texts share the idea that while we may always continue to strive for greater virtue, a commitment to this end “holds for God as possession” (AA 5:123n, cf. AA 6:67). ‘Grace’, in this sense, may either be understood as “perspectival” (Palmquist 2010), i.e., based upon God's judging us as “a perfected whole” (6:67), or entailing a judgment “which do[es] not harmonize with justice” (AA 5:123) since God takes the lesser condition of endless progress as tantamount to our having reached the goal (AA 5:123n).

Third, unlike the Second *Critique* where Kant is not as diligent about distinguishing the particular doctrinal issues with which the Highest Good is intertwined, in *Religion* he explicitly attends to the Christian distinction between Sanctifying and Justifying Grace. Both have to do with our ‘fallen’ condition and with what God does for us in light of that condition. Justifying grace pertains to the “debt of sin,” which, according to the Augustinian tradition, we have inherited from Adam as the result of his “primal” act of disobedience. The Crucifixion then repays this debt such that, through the doctrine of Vicarious Atonement, one can partake of a “foreign satisfying merit.” Justifying Grace, in short, consists in the repayment of our original debt.

In the Second *Critique*, we can see traces of this doctrinal issue: even with our endless effort, we still cannot become *fully* worthy of happiness, a shortcoming that leads God to grant us an “indulgence.” This matter is then treated with more precision in *Religion*, where Kant proposes that the debt of sin is neither forgiven (since this would violate divine justice – AA 6:73, 6:76) nor repaid (since the debt is infinite – AA 6:72), but is rather “removed” or unstruck [*entschlagen*] (AA 6:76). Thus, in agreement with the Augustinian tradition, Kant believes that there is no “justification through works.”<sup>6</sup> But he parts ways with Augustinians on the doctrine of vicarious atonement: for Kant, moral liabilities are not “transmissible” (AA 6:72) and radical evil is the most personal of debts. Hence, as Mariña (1997) has previously argued, Kant diverges at this juncture from the Calvinist and orthodox Lutheran notion of “forensic” justification, according to which justification is no more than an edict by God. Moreover, he also rejects the economic tropes of Reformed theology (Clem 2018), opposing the idea that sin is to be modeled as a

---

<sup>6</sup> It is important to note that Kant does *not* follow the Augustinian tradition on an essential point, namely, that we cannot earn any moral merit through our efforts. Rather, it is essential to Kant's moral religion that we become “well-pleasing to God” through “good life conduct” (AA 6:171). The logic of the doctrine of Justification is distinct from this and is rather based upon the antecedent claim that we are born with an infinite debt to God. Kant thus divides up in Part II the question of sanctification (the “first difficulty”) from the question of justification (the “third difficulty”). In fact, the analysis Kant is developing in the three “difficulties” particularly concerns the debate between the Lutheran Pietist treatment of Justification/Sanctification and the orthodox Lutheran position. Space does not allow us to work through the details here. See Clem 2018.

“financial debt” (AA 6:72). Hence, while Kant engages with the doctrine of Justifying Grace and advances a view on Justification which conforms to rational religion, that view is developed through a critique of the standard theological positions of his day.

Sanctifying Grace, much more so than Justification, is the focus of this dossier, and to it we now turn. This kind of grace concerns the role assigned to God in our transition from evil to good. But before we discuss this point directly, one first needs to understand the underlying theological and moral problematics.

Calvinists and orthodox Lutherans describe our fallen condition as one of “Total Depravity.” They hold that Adam’s ‘primal sin’ changed human nature such that we are now “slaves to sin,” consumed by self-interest, confused about the proper order of “lesser” and “greater” goods, and in our “fallen” state, incapable of any moral improvement. While there are some nuances which distinguish Calvin and Luther from Augustine’s own views, all these thinkers share a *privation* account of moral evil, according to which our “fallen” condition is understood in terms of cognitive and volitional *incapacities* that prevent us from making correct moral judgments and willing the good.

The broad strokes of the Augustinian tradition are explored by Kant through the main body of Part One. That is, as we mentioned earlier, throughout the *Religion* Kant forges a path through a series of Christian doctrines, introducing and then testing their various tenets. Part One tenders for review the Augustinian doctrine that humanity is evil, considers the various claims bound up with this doctrine (innateness, inheritance, corruption of our powers, etc.), and examines these doctrinal claims in light of the moral principles of “genuine” religion.

For example, Kant recognizes that if humanity is *universally* evil, then there must be something about us “*qua* human” that makes this so—and hence evil must be innate. However, the notion of an innate moral condition is for Kant inadmissible if it suggests that this condition is *not* a product of our wills. Hence, he rejects various conventional positions (biological inheritance, determining inclination, and other “natural causes” (AA 6:21)), because the only way to reconcile innateness with the “predicates *morally* good and *morally* evil” (AA 6:21) is by taking it “only *in the sense* that [such character] is posited as the ground antecedent to every use of freedom given in experience” (AA 6:22).

Similar considerations lead Kant to *reject* common doctrinal views on moral evil according to which, in our “fallen” condition, our volitional powers are so deeply compromised that we are enslaved to our “sensuous nature” (AA 6:35) and/or that there is a total “corruption of [our] morally legislative reason” (AA 6:35). Hence, Kant rejects the dominant privation model of evil (corruption of our faculties) and kindred views, and opts instead for explaining “the difference, whether the human being is good or evil,” in terms of the order of incentives in one’s “supreme maxim.”

As for how Sanctifying Grace relates to the above, let us begin with the Pauline distinction between the “old man” and the “new man,” a distinction Kant often uses both in the *Religion* and the *Conflict of the Faculties*. Given the privation model of the Augustinian tradition, the “old man” is a “slave to sin” and lacks the capacities needed to



become a “new man,” i.e., one who is able to know and will the good. Accordingly, this tradition holds that our moral transformation is only possible through divine aid. For Augustine and his Reformation followers, there are a select few, the “elect,” who have been restored, but such restoration is completely independent of anything they have done to help bring it about. As we can neither know the good nor will it, in our fallen condition we cannot even *ask* for help. Instead, every human being moves forward in life governed solely by self-interest and confused about what the good really is. The “elect,” on the other hand, have their capacities spontaneously restored to them, and, once restored, they enjoy an ongoing *concursus* or cooperation by which they continue to persevere in the good, exercising and strengthening their newly acquired powers of moral judgment, resisting self-interest, and choosing the good.

We may thus distinguish between three traditional theological positions regarding our moral transformation. First, there are those who believe that we are ‘slaves to sin’, utterly unable, unwilling to embrace, and ignorant of, the moral alternative. The more stringent brands of Augustinianism, including Calvinists and orthodox Lutherans, subscribe to this view. They hold that the “elect” are “passive” recipients of sanctifying grace and can do nothing to usher it in. They also hold that while the “old man” could in theory resist divine aid, grace is nonetheless “efficacious” and resistance is futile. Second, there are qualified versions of this stringent view, which hold that we cannot positively act towards the good but can nonetheless resist sin and thus move towards “quiescence.” Aquinas, for example, is taken by some to hold this position (Stump 2001). Finally, and this is the main Kantian target, there are those who hold that while we can do nothing to *merit* aid, God has already granted us a “prevenient” power which, despite our fallen condition, preserves within us the ability to ask for help. This view holds that prevenient grace provides enough of a flicker of freedom to allow us then to request divine assistance.

While this third view of Sanctifying Grace has gained in popularity over the centuries, it was harshly attacked by Calvinists and orthodox Lutherans up through Kant’s time. It is usually associated with Jacobus Arminius, who, though condemned by the Calvinist Synod of Dort, was received with more favor among Catholics that found him compatible with the Council of Trent (Session 6, Chapter 5). Arminius contributed to the resurgence of Moravianism, laid the foundations for Methodism, and paved the way for numerous further Protestant movements, including Lutheran Pietism.<sup>7</sup>

As we will discuss in the next section, a number of contemporary interpretative disputes regarding *Religion* arose because various Reformed philosophers of religion wanted to read the text through a Calvinist lens. That is, just as secular readers of the

---

<sup>7</sup> In his introduction to the Cambridge paperback edition of Kant’s *Religion*, Robert Adams distinguishes between prevenient, sanctifying and justifying grace. Our presentation of these three types of grace roughly track with Adams’, though it is far from clear that he is correct that there is “no place in the Kantian scheme of things for *prevenient* grace” (Adams 1998, p. xxi). Its role would be different from its use in Arminianism and Pietism, but we regard this as a minor quibble. We do, however, want to indicate that Adams may give the wrong impression about the historical place of prevenient grace. While it was recognized by Augustine as well, it gains a particular importance through and after Arminius as a response to the doctrine of election and as a qualification to the Reformation doctrine of Total Depravity.



*Religion* do not usually attend to its theological background, theologians of various orthodoxies fail to recognize that Lutheran Pietism is the principal tradition that Kant has in mind when discussing Sanctifying Grace, both in the *Religion* and the *Conflict of the Faculties*.

Given Kant's comparative/evaluative procedure, it should be apparent why he could not possibly endorse the more stringent Augustinian view: its doctrines of "Total Depravity" and the utter impotence of the passive recipients of a transformative grace do not square with the demands of moral agency. But Kant also opposes the more moderate version that grants Sanctifying Grace to those who merely "ask for it" (AA 6:51), for they still share a fundamental Augustinian assumption, namely, that our moral transformation does not lie within our own powers. We may (due to Prevenient Grace) remain capable of asking for it, but a transformation through a mere request falls well short of the moral significance Kant places on our own efforts.

Kant launches an outright invective against this position at the end of Part One, where he proposes that the doctrinal spectrum, from total passivity to merely "asking" for help, is an expression of our moral corruption, a symptom of rather than a cure for the propensity to evil. As he complains, those who find "moral labor vexing" have conjured up "under the pretext of natural impotence" doctrines through which they can be relieved from the "expectation of self-improvement" (AA 6:51). Thus, what Kant is indicating is that the broader Augustinian tradition, from the more stringent Calvinists and orthodox Lutherans to the gentler Pietists and Moravians, yields to the temptation arising from the first of the three propensities to evil, frailty, i.e., the self-imposed delusion that we are incapable of willing the good.

In light of this diagnosis, Kant concludes that all theological variants of this conception promote a "religion of roagation (of mere cult)" (AA 6:51). Instead of encouraging us to engage in the hard work of moral improvement, they tell us that we need merely "ask" for grace—a plea that "amount[s] in fact to doing nothing" (AA 6:51). In contrast to this 'lazy' attitude, moral religion demands that we be responsible for our own moral standing, the authors of our own character.

The problem Kant still faces, however, is that, having given priority to self-interest over morality, the logic of radical evil makes any attempt to reverse the current volitional order seem impossible. For, once the inversion has taken place, self-interest could only perpetuate itself. Hence, even if the moral predisposition has not lost any of its influence on our will, there is still the question of how, once we have given priority to self-interest, could we ever choose to subsequently subordinate our self-interest to morality.

Yet, the Kantian thesis of 'ought implies can' tells us that the release from evil must be possible. Furthermore, Kant's commitment to transcendental freedom entails that our supreme maxim is not chosen once and for all, but must instead be constantly reaffirmed through an ongoing choice, counted at each moment as if it were always "an *original* exercise" of the will (AA 6:41). No matter how self-reinforcing and self-

perpetuating the logic of radical evil might be, the demand to undergo a Change of Heart remains unabated.

The question of Sanctifying Grace arises in response to this problem. Kant does not tell us that divine aid is necessary, nor does he tell us that it is not. For even if we must exercise our powers to become a 'new man', this does not mean that the moral transformation occurs through our powers *alone*. Instead, and in keeping with the point of the *parerga* (which mark the limits of what rational religion can determine), 'ought implies can' commits us to the belief that the transformation is possible, but leaves open the question of how the change might happen. All a religion that remains within the boundaries of reason can acknowledge is that "[E]veryone must do as much as it is in his power to do," and only then "can he hope that what does not lie in his power will be made good by cooperation from above" (AA 6:52).

### §3. Genealogy of an Interpretative Impasse

The history of interpretation of *Religion* has, unfortunately, failed to understand the precise status of this hope. We believe that this is due, in large measure, to Karl Barth's 1947 *Protestant Theology in the Nineteenth Century*. In this book, Barth is skeptical about Kant's grasp of Reformation theology (Barth 1947, p. 285) and believes that his own philosophical biases stood in the way of a proper understanding of the Christian Story (Barth 1947, p. 293). Barth further describes Kant's philosophical theology as vitiated by an internal contradiction: while Part One of *Religion* follows an Augustinian account of moral evil, Kant then slides into a "vulgar Pelagian" account of grace (Barth 1947, p. 289).

With few changes, this picture still dominates the current state of the debate: interpreters like Philip Quinn, Nicholas Wolterstorff, John Hare and Gordon Michalson have suggested that, despite Kant's considerable philosophical genius, he was out of his depth in the *Religion* and failed to manage the complexities of Christian theology. Wolterstorff, for example, describes the *Religion* as being "riddled with irrationalities," torn by an "internal contradiction" between an Augustinian conception of Original Sin and a so-called "Stoic maxim," which holds that "a person's moral worth is determined entirely by that person himself" (Wolterstorff 1991, p. 48). Similarly, Michalson claims that the *Religion*'s elaborate "explanatory apparatus" is engendered by conflicting Christian and Enlightenment "habits of mind," and, far from reconciling the two, Kant instead "cannot fully understand his own position" (Michalson 1990, p. 9).

Numerous Kant specialists have (perhaps without realizing it) accepted the Barthian interpretation. They find an unstable compromise at the heart of *Religion*: while the main body of Part One expresses Kant's Augustinian commitments, after the "General Remark" Kant retreats to the model of moral agency we find in his other writings, according to which "only that can be morally good which can be attributed to us as performed by ourselves" (Barth 1947, p. 284). Hence, instead of coming to Kant's defense, many influential Kantians have granted the premises of the Barthian critique: Wood, for example, seeks to explain away the Augustinianism of Part One as residual

“nostalgia towards” pre-modern inquiries into the nature of morality (Wood 2005), and Allison accepts that Kant’s account of evil is “closer to that of the later Augustine” than either “the standard views of the Enlightenment” or the “intuitions of present day philosophers” (Allison 2002, p. 338).

The interpretative consensus is thus that Kant is at war with himself and presents various Christian theses that undermine his core moral principles. The more militant approaches give up all pretense of reconciliation and embrace one extreme or another: some see Kant as a champion of Christian orthodoxy (Firestone/Jacobs 2008), and others see him as refusing to engage any kind of “theological speculation” whatsoever (DiCenso 2012, p. 117). In one way or another, the majority of scholarship grants that there are “conundrums” and “internal contradictions,” allowing the Barthian objections to the *Religion* to continue to reign.

We propose, however that these views are neither accurate to the text nor needed: the legacy of “conundrums” and “wobbles” owes more to our failing to grasp the basic theology than to Kant’s not “fully understand[ing] his own position.”

This is so, first and foremost, because the majority of contemporary interpreters (from Barth through Quinn, Wolterstorff, Hare, all the way to Firestone/Jacobs) read the *Religion* through a Calvinist lens. Hence, passages where Kant considers an economy of salvation which diverges from the doctrines of election and the passive reception of grace are either seen as indicative of his own lack of theological competence or, with Firestone/Jacobs, made to fit Calvinist orthodoxies. But this interpretation is both misleading and misguided. As we have seen, the primary focus of *Religion* is Pietism, which resembled not the theology of Calvinism but much more that of its arch-rival, Arminianism. Like the Arminians, and unlike the Calvinists (and other orthodox Lutherans), Pietists denied total depravity and accepted that we were still able to at least “ask” for grace. Once we recognize this fact, we begin to see that Kant was not out of his theological depth—he was navigating other waters altogether.

Moreover, the Calvinist bias of the Barthian framework is also to blame for the impression of a “conundrum” or “wobble” at the heart of Kant’s position. As explained in the preceding sections, Kant’s aim is to compare rational religion and historical faith in order to determine how much “unity” or “compatibility” there is between them. What seems to be at best meandering, and at worst a litany of “wobbles,” is instead an invitation to identify and assess particular doctrinal tenets in terms of their moral implications. Seen this way, Kant does not stumble through a maze of Christian doctrines, but hovers above them, comparing and evaluating each against the principles of rational religion.

Consider, once again, the widespread claim that Part One reflects an Augustinian conception of evil. If one examines Part One in relation to the particulars of this conception, a very different picture emerges. For, quite in contrast with the Augustinian tradition, Kant maintains that our moral condition is not acquired in time (by Adam’s Primal Sin – AA 6:43), nor is it inherited from parent to child (AA 6:40) but must be “freely chosen” (AA 6:21). Furthermore, evil does not involve a “corruption of the morally

legislative reason" (AA 6:35), nor does it entail a loss of freedom (AA 6:41). Hence, while the Augustinian advances a privation account of evil, Kant explicitly denies this view: evil is not a matter of corrupt faculties, but of an ethical inversion of the order of incentives, a reversal that bespeaks the presence of an "active and opposing cause" (AA 6:57), a "positive principle" (AA 6:59) within us.

Accordingly, Part One is *not* "firmly ensconced within the Augustinian tradition" (Quinn 1988, p. 91). Nor is Kant "accommodating" it in any fashion (Wood 1970, p. 246), nor is he building it around tenets "fundamental to Augustine" (Beiser 2006, p. 594), nor resting upon an "Augustinian metaphysic" (Firestone and Jacobs 2008, p. 136). Kant's account of evil is *not* at all "closer to that of the later Augustine" than "the standard views of the Enlightenment" (Allison 2002, p. 338). Rather, the "Augustinian echoes" (Mariña 1997, p. 379) of Part One are not "wobbles," but doctrinal tenets tendered for evaluation which often fail to pass the moral test Kant applies to them.

In short, if we read *Religion* according to Kant's own comparative/evaluative intent, the wobbles and failures are revealed for what they are: a byproduct of a misguided interpretative legacy, an external imposition that prevents us from reading the text in its own terms. Kant can be said to have misunderstood Reformation theology—but only if one expects to find pure Calvin in it. Once we abandon this old prejudice, the whole conundrumist apparatus (the dichotomy between Augustinianism and Pelagianism) collapses with it. There is really no "wobbling" between Enlightenment values and Christian orthodoxy, for Kant is not oscillating between two worlds, but engaged in the project of assessing the "unity" or "compatibility" of Christianity and moral/rational religion.

#### §4. New Voices

The APA symposium was motivated by the will to overcome the unfortunate legacy of the Barthian paradigm. In "Why Is Kant Noncommittal about Grace?", Robert Gressis questions the fundamental assumption of the conundrumist reading, namely, that there is an outright contradiction between our duty to make ourselves who we are (from the moral point of view) and the need of God's help in becoming good. This dichotomy presupposes that dependence on another and personal responsibility are contradictory notions—and hence ignores the many instances in which our agency goes hand-in-hand with our receptivity to help. This is the case, for instance, in student/teacher relationships, where *responsiveness* to the teacher's guidance does not preclude, but rather bolsters, our agency in learning. Once the appearance of contradiction disappears, Gressis believes, one can appreciate what is truly puzzling in the Kantian account, namely, that Kant neither asserts nor denies the need for divine aid—he leaves the question open, and there are good reasons for that. For, to the extent that settling the issue of grace *theoretically* is for Kant impossible, any dogmatic position (affirming or denying it) is equally untenable. The question, thus, must be decided on strictly practical grounds.

At this level, Gressis argues, it makes sense for Kant to be noncommittal about grace, since such position avoids all sorts of vices. While certainty about the availability of divine aid would promote either laziness or arrogance, tempting the recipient to abdicate all responsibility for her moral situation or to fancy herself to be divinely chosen, certainty about its absence would convince her of the futility of any effort, or, if she were to undertake the task of moral regeneration against all odds, generate an inflated sense of her own powers. Although the dangers of passivity, smugness, and self-conceit, Gressis contends, are all connected to the radicalism of evil, Gressis is careful to avoid what he calls the “logical interpretation” of the *Gesinnung*, according to which it is as impossible for an evil person to act against her self-love as it is for a good person to act immorally. Both extremes contradict our moral experience, and overlook the role that fortuity and exemplary moral action can have in restoring our moral predisposition to its rightful place.

Dennis Vanden Auweele engages in a similar project of redescription of the conundrum problematic. In “Kantian Grace as Ethical Gymnastics,” he sidesteps the Barthian framework and proposes a different genealogy, one that traces the source of the scholarly confusion to a contemporary of Kant, the theologian Gottlob Christian Storr. Storr, in Vanden Auweele story, is the father of a long string of reductionistic readings of *Religion*, according to which the Kantian project of reconciling faith and reason threatens both what is distinctively religious in Christianity and what is genuinely philosophical in rationalism. This perception underlies much of the 19<sup>th</sup> century reception of Kant’s philosophy of religion (Schopenhauer and Hegel are cases in point) and prevents us from reading *Religion* in Kant’s own terms. If we do so, Vanden Auweele argues, we will realize that the “overall project in *Religion* is twofold, namely to provide a transcendental deduction of the elements of pure rational religion and to test whether or not Christian religion is in tune with these elements” (288). So construed, the relation between faith and reason is not so much one of tension; it is, instead, one of mutual reinforcement and complementarity. The appeals to the “irrational or hyperrational” aspects of Christianity are not a failure, but corollary of an “unavoidable limitation of human reason” (AA 6:115).

Although at times Kant laments this limitation and hopes that human beings will kick away the ladder of historical faith, he also realizes that the allegories and narrative structures of Christianity play an essential role in our moral education. For, in so far as the moral law presents itself in the form of the categorical imperative, pure practical reason presupposes a whole aesthetic dimension that allows it to gain “access to the human mind and [have] influence on its maxims” (AA 5:151). Historical faith, therefore, is instrumental in making “objectively practical reason *subjectively* practical as well” (ibid.)—it belongs, therefore, to the ‘second part of ethics’, the impure underbelly of morality without which purity would never take a hold. To make his case, Vanden Auweele analyzes the “antinomy of faith” in *Religion* III, where Kant dissolves the conflict between historical faith and moral religion by reminding us that the son of God is a moral archetype that resides within pure practical reason itself, and hence is “all the same whether I start out from it (as *rational* faith) or from the principle of good life conduct” (AA 6:119).

In “Kant’s Robust Theory of Grace,” Jaqueline Mariña presents a different argumentative strategy: instead of redescribing the conundrumist dilemma (as the other panelists do), Mariña redescribes the different meanings of grace in Kant’s *Religion*. The inconsistencies surrounding the overcoming of the problem of evil disappear, she argues, “once it is recognized that Kant works with three understandings of grace. These are: a) grace and the God within, b) grace and the transformation of the fundamental orientation, and c) grace that can be laid hold of” (303). This redescription is meant to elude many of the problems associated with the standard understanding of grace in the Lutheran tradition (which plague the second of the three meanings she proposes), and direct our attention instead to the possibility of divine aid in the historical arena (c) and the sacred character of the moral law (a). The latter leads Kant to overhaul the Leibnizian distinction between nature and grace, for the divinity of the moral law is no longer seen as an alien force, but is internalized and gives rise to what Mariña calls our “*graced nature*”, a ‘nature’ that is epitomized by the prototype (*Urbild*) of moral perfection. This conception, she argues, allows Kant to bypass the problems associated with the ‘Total Depravity’ tradition. Seen this way, grace is not a response to the doctrine of radical evil—“the problem of radical evil”, instead, “must itself be understood in terms of an outright refusal to allow the [moral] predisposition and the archetype of the Son of God (the Christ within) to be effective in us” (306). Such an interpretation upends common wisdom: the relation to God is not a consequence of—but rather precedes—our fall into evil.

Similarly, the kind of divine aid that “must be laid hold of” avoids the problems that lead conundrumist writers to attribute to Kant a heteronomous conception of grace. Mariña traces back this type of faith to the “hope that what lies outside [our] power will be supplemented by the *supreme wisdom in some way or another*” (AA 6:171), for such hope is only available to those who have *already* undergone a fundamental change of disposition and helps them carry on their moral struggle. This help, Kant tells us, presupposes acceptance from the agent, i.e., “he must incorporate this positive increase of force into his maxim” (AA 6:44) and is hence fully *active in receiving it*. The complex relation between natural religion and revelation, Mariña argues, is also a case in point: in order for the latter to speed up and increase our awareness of the moral core of religion, human beings must properly receive it, allowing the revelation to “spark the moral imagination” and make the moral demands a “point of focus for [...] reflection and engagement” (314). A further example is the founding of the ethical community: even if God is needed, we must act as if everything depended on us if we are to “hope that a higher wisdom will provide the fulfillment of [our] well-intentioned effort” (6:101). Thus, Mariña concludes, even though Kant is skeptical about the second type of grace, for in the process of moral transformation it leads us to abdicate our responsibility for the change, there is enough substance in the two other conceptions of grace to yield a “remarkably positive” (i.e., “robust”) understanding of [divine aid] in Kantian philosophy (304).

## Conclusion

A guiding assumption of this guide has been that the maligned and misunderstood character of Kant's *Religion* is not so much a reflection of its forbidding nature, but rather of our own prejudices and limitations in understanding its theological underpinnings. We contemporary philosophers live in a world in which the religious debates that occupied Kant seem alien or quaint, remnants of a pre-modern past. They are not, as William James would put it, live options for us. The goal of our introduction has not been to restore them to life, but to return the reader to the concerns of Kant's own day and to understand the *Religion* as a work quite purposefully and consciously written for his own contemporaries, people who lived amidst two clashing worldviews and were in search of a path forward.

## Bibliography

- Allison, Henry E. 2002. "On the Very Idea of a Propensity to Evil." *Journal of Value Inquiry* 36: 337-348.
- Barth, Karl. 1947. *Die protestantische Theologie im 19. Jahrhundert*. Zürich: Theologischer Verlag translated as Protestant Theology in the Nineteenth Century. Eerdmans 2002.
- Beiser, Frederick. 2006. "Moral Faith and the Highest Good." In *The Cambridge Companion to Kant and Modern Philosophy*. Paul Guyer, (ed). Cambridge: Cambridge University Press, 588-629.
- Clem, Stewart. 2018. "Dropping the Debt: a new Conundrum in Kant's Rational Religion." *Religious Studies*, in press.
- DiCenso, James. 2012. *Kant's Religion within the Boundaries of Mere Reason: A Commentary*. Cambridge: Cambridge University Press.
- Firestone, Chris L., and Nathan Jacobs. 2008. *In Defense of Kant's Religion*. Bloomington: Indiana University Press.
- Frierson, Patrick R. (2010) "Kantian Moral Pessimism." In S. Anderson-Gold and P. Muchnik (eds.), *Kant's Anatomy of Evil*, Cambridge: Cambridge University Press, 33-56.
- Hare, John. E. 1996. *The Moral Gap: Kantian Ethics, Human Limits, and God's Assistance*. Oxford: Clarendon Press.
- (2017). "Kant and Theology." *Toronto Journal of Theology* 33:1, 69-73.
- Mariña, Jacqueline. 1997. "Kant on Grace: A Reply to His Critics." *Religious Studies* 33:4, 379-400.
- Michalson, Gordon E. 1990. *Fallen Freedom: Kant on Radical Evil and Moral Regeneration*. Cambridge: Cambridge University Press.
- Palmquist, Stephen. 2000. *Kant's Critical Religion*. London: Ashgate.
- (2010) "Kant's Ethics of Grace: Perspectival Solutions to the Moral Difficulties with Divine Assistance." *Journal of Religion* 90(4): 530-53.



- Pasternack, Lawrence. 2014. *Kant on Religion within the Boundaries of Mere Reason: an Interpretation and Defense*. London: Routledge.
- Pasternack, Lawrence. 2017. "The 'Two Experiments' of Kant's Religion: Dismantling the Conundrum." *Kantian Review*, 22:1, 107-131.
- Quinn, Philip L. 1988. "In Adam's Fall, We Sinned All." *Philosophical Topics* 16:2, 89-118.
- Stump. 2001. "Augustine on Free Will." *The Cambridge Companion to Augustine*. Cambridge University Press, 124-147.
- Velkley, Richard. 2014. "Culture and the Limits of Practical Reason." *Kant's Religion within the Boundaries of Mere Reason: A Critical Guide*. Cambridge: Cambridge University Press, 233-249.
- Wolterstorff, Nicholas. 1991. "Conundrums in Kant's Rational Religion." *Kant's Philosophy of Religion Reconsidered*, Rossi, Philip and Wreen, Michael (eds). Bloomington: Indiana University Press.
- Wood, Allen W. 1970. *Kant's Moral Religion*. Ithaca: Cornell University Press.
- Wood, Allen W. 2005. "Kant's History of Ethics." *Studies in the History of Ethics* vol. 1.

